Biometrika (2018), **103**, 1, *pp.* 1–11
Printed in Great Britain

Supplementary Material for "Generalized Meta-Analysis for Multiple Regression Models Across Studies with Disparate Covariate Information"

BY PROSENJIT KUNDU
Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland, 21205, U.S.A.
pkundu@jhu.edu

RUNLONG TANG
Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland, 21205, U.S.A.
rtang15@jhu.edu

AND NILANJAN CHATTERJEE
Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland, 21205, U.S.A.; Department of Oncology, School of Medicine, Johns Hopkins University, Baltimore, MD, 21205, U.S.A.
nchatte2@jhu.edu

1. ASYMPTOTIC EQUIVALENCE OF GENMETA ESTIMATOR AND SIMPLE META-ANALYSIS ESTIMATOR WHEN ALL THE REDUCED MODELS ARE THE SAME TO THE MAXIMAL MODEL

When all the reduced models are the same to the maximal model, it follows $\theta_k^* = \beta^*$, $X_{A_k} = X$ and $g_k = f$ for $k = 1, 2, \ldots, K$. Then, for each k, $u_k(X; \beta^*, \theta_k^*) = u_k(X; \beta^*, \beta^*) = \int s_k(y \mid X_{A_k}; \beta^*) f(y \mid X; \beta^*) dy = 0$. By the definition of Δ, we have $\Delta = 0$. On the other hand, assuming $E_{Y|X}\{\nabla_{\theta_k} s_k(\theta_k^*)\} = \nabla_{\theta_k} E_{Y|X}\{s_k(\theta_k^*)\}$ with $s_k(\theta_k^*) = s_k(Y \mid X_{A_k}; \theta_k^*)$, it follows $\Lambda_k = (1/c_k) I(\theta_k^*)$, where $I(\theta_k^*)$ is the Fisher's information matrix of g_k or f. Then, the optimal C is

$$C_{\text{opt}} = \Lambda^{-1} = \text{diag}(c_1\Sigma, \ldots, c_K\Sigma),$$

where $\Sigma = I(\theta_k^*)^{-1}$. Denote as $\hat{C}_{\text{opt}}$ a consistent estimator of C_{opt}. Then, the GENMETA estimator with $\hat{C}_{\text{opt}}$ is

$$\hat{\beta}_{\text{opt}} = \text{argmin}_{\beta} U_n^T(\beta, \hat{\theta}) \hat{C}_{\text{opt}} U_n(\beta, \hat{\theta}).$$

Under regularity conditions similar to those in Theorem 1, $\hat{\beta}_{\text{opt}} \to \beta^*$ in probability. By Mean Value Theorem,

$$U_n(\hat{\beta}_{\text{opt}}, \hat{\theta}) = U_n(\beta^*, \hat{\theta}) + G_n(\bar{\beta}, \hat{\theta})(\hat{\beta}_{\text{opt}} - \beta^*), \tag{1}$$

where $\bar{\beta}$ is the mean value and $G_n(\bar{\beta},\hat{\theta}) = \partial U_n(\beta,\hat{\theta})/\partial\beta \mid_{\beta=\bar{\beta}}$. By the first order condition, $\hat{\beta}_{\text{opt}}$ satisfies $G_n^T(\hat{\beta}_{\text{opt}},\hat{\theta})\hat{C}_{\text{opt}}U_n(\hat{\beta}_{\text{opt}},\hat{\theta}) = 0$. Left-multiplying (1) by $G_n^T(\hat{\beta}_{\text{opt}},\hat{\theta})\hat{C}_{\text{opt}}$, it follows

$$\hat{\beta}_{\text{opt}} - \beta^* = -\{G_n^T(\hat{\beta}_{\text{opt}},\hat{\theta})\hat{C}_{\text{opt}}G_n(\bar{\beta},\hat{\theta})\}^{-1}\{G_n^T(\hat{\beta}_{\text{opt}},\hat{\theta})\hat{C}_{\text{opt}}U_n(\beta^*,\hat{\theta})\} \tag{2}$$

Also,

$$G_n(\hat{\beta}_{\text{opt}},\hat{\theta}) = \frac{\partial}{\partial\beta}U_n(\beta,\hat{\theta}) \mid_{\beta=\hat{\beta}_{\text{opt}}} = \begin{pmatrix} \frac{\partial}{\partial\beta}u_1(\beta,\hat{\theta}_1) \mid_{\beta=\hat{\beta}_{\text{opt}}} \\ \vdots \\ \frac{\partial}{\partial\beta}u_K(\beta,\hat{\theta}_K) \mid_{\beta=\hat{\beta}_{\text{opt}}} \end{pmatrix}.$$

Under regularity conditions similar to those in Theorem 1, $\partial u_k(\beta,\hat{\theta}_k)/\partial\beta \mid_{\beta=\hat{\beta}_{\text{opt}}} = \Sigma^{-1} + o_p(1)$ for each k. Then,

$$G_n(\hat{\beta}_{\text{opt}},\hat{\theta}) = \begin{pmatrix} \Sigma^{-1} \\ \vdots \\ \Sigma^{-1} \end{pmatrix} + o_p(1). \tag{3}$$

Similarly,

$$G_n(\bar{\beta},\hat{\theta}) = \begin{pmatrix} \Sigma^{-1} \\ \vdots \\ \Sigma^{-1} \end{pmatrix} + o_p(1). \tag{4}$$

On the other hand, under regularity conditions similar to those in Theorem 1, $u_k(\beta^*,\hat{\theta}_k) = -\Sigma^{-1}(\hat{\theta}_k - \beta)^* + o_p(1/n^{1/2})$. Then,

$$U_n(\beta^*,\hat{\theta}) = -\begin{pmatrix} \Sigma^{-1}(\hat{\theta}_1 - \beta^*) \\ \vdots \\ \Sigma^{-1}(\hat{\theta}_K - \beta^*) \end{pmatrix} + o_p(1/n^{1/2}). \tag{5}$$

Hence, by (2), (3), (4), (5) and Slutsky's theorem,

$$\hat{\beta}_{\text{opt}} - \beta^* = \Big(\sum_{k=1}^K c_k\Big)^{-1}\Big\{\sum_{k=1}^K c_k(\hat{\theta}_k - \beta^*)\Big\} + o_p(1/n^{1/2}). \tag{6}$$

On the other hand,

$$\begin{aligned} \hat{\beta}_{\text{meta}} - \beta^* &= \Big\{\sum_{k=1}^K \Big(\frac{\hat{\Sigma}_k}{n_k}\Big)^{-1}\Big\}^{-1}\Big\{\sum_{k=1}^K \Big(\frac{\hat{\Sigma}_k}{n_k}\Big)^{-1}\hat{\theta}_k\Big\} - \beta^* \\ &= \Big(\sum_{k=1}^K c_k\Big)^{-1}\Big\{\sum_{k=1}^K c_k(\hat{\theta}_k - \beta^*)\Big\} + o_p(1/n^{1/2}). \end{aligned} \tag{7}$$

Therefore, by (6) and (7), $\hat{\beta}_{\text{opt}} = \hat{\beta}_{\text{meta}} + o_p(1/n^{1/2})$.

2. Newton-Raphson's Method and Iteratively Reweighted Least Squares Algorithm

In this section we provide a derivation of the Newton-Raphson's method for GENMETA with generalized linear models. As in Section 2.3, we assume that the maximal and reduced models belong to the class of GLM (McCullagh & Nelder, 1989). Specifically, assume the densities of $Y \mid X$ and $Y \mid X_{A_k}$ are of the forms

$$f(y \mid x; \beta, \phi) = \exp(\{1/a(\phi)\}(yh(x^T\beta) - b\{h(x^T\beta)\}) + c(y;\phi)),$$

and

$$g_k(y \mid x_{A_k}; \theta_k) = \exp(\{1/a(\phi_k)\}(yh(x_{A_k}^T\theta_k) - b\{h(x_{A_k}^T\theta_k)\}) + c(y;\phi_k)),$$

respectively, where $a(\cdot)$, $b(\cdot)$ and $c(\cdot)$ are known functions, $h(\cdot) = b'^{-1}(g^{-1}(\cdot))$, g is a monotone and differentiable link function, and ϕ and ϕ_k are the dispersion parameters of the maximal and the kth reduced models, respectively. Recall that we assume the maximal and the reduced models have the same link function g. However, both the GENMETA and the Newton-Raphson's method are flexible to allow the maximal and the reduced models to have different link functions. We also assume $X = \cup_{k=1}^{K} X_{A_k}$, where the vectors of the covariates are viewed as sets without confusion. Denote the dimensions of θ_k and β as d_k and p, respectively. Assume $d = \sum_{k=1}^{K} d_k \geq p$ since the parameters of the maximal model will not be identifiable if $d < p$.

2.1. *Case I : ϕ and ϕ_k's are known.*

The log-likelihood of g_k is

$$l_k(y \mid x_{A_k}; \theta_k) = \{1/a(\phi_k)\}(yh(x_{A_k}^T\theta_k) - b\{h(x_{A_k}^T\theta_k)\}) + c(y;\phi_k).$$

Then, the score function is

$$s_k(y \mid x_{A_k}; \theta_k) = \{1/a(\phi_k)\}\{y - g^{-1}(x_{A_k}^T\theta_k)\}h'(x_{A_k}^T\theta_k)x_{A_k}.$$

Then,

$$u_k(x;\beta,\theta_k) = E_{Y|X}s_k\{(y \mid x_{A_k};\theta_k)\} = \{1/a(\phi_k)\}\{g^{-1}(x^T\beta) - g^{-1}(x_{A_k}^T\theta_k)\}h'(x_{A_k}^T\theta_k)x_{A_k}.$$

Thus, the vector of empirical moment functions for β is

$$U_n(\beta) = P_n \begin{pmatrix} u_k(X;\beta,\hat{\theta}_k) \\ u_k(X;\beta,\hat{\theta}_k) \\ \vdots \\ u_k(X;\beta,\hat{\theta}_k) \end{pmatrix},$$

where P_n is the empirical measure with respect to the reference sample.

Let $Q_n(\beta) = U_n^T(\beta)CU_n(\beta)$ where C is a $d \times d$ positive definite matrix. The goal is to find the minimizer of $Q_n(\beta)$. Its equivalent to solving the equation

$$D_n(\beta) = 0,$$

where $D_n(\beta) = G_n^T(\beta)CU_n(\beta)$ and $G_n(\beta) = \partial U_n(\beta)/\partial\beta$ is a $d \times p$ matrix. Then, the tth iteration step for the Newton-Raphson's method is

$$\beta^{(t+1)} = \beta^{(t)} - J_n(\beta^{(t)})^{-1}D_n(\beta^{(t)}), \tag{8}$$

where $J_n(\beta) = \partial D_n(\beta)/\partial\beta$ is a $p \times p$ matrix.

Next, we write $D_n(\beta)$ in a matrix form. The matrix form of $G_n(\beta)$ is

$$G_n(\beta) = P_n \begin{pmatrix} [a(\phi_1)g'\{g^{-1}(X^T\beta)\}]^{-1}h'(X_{A_1}^T\hat{\theta}_1)X_{A_1}X^T \\ \vdots \\ [a(\phi_K)g'\{g^{-1}(X^T\beta)\}]^{-1}h'(X_{A_K}^T\hat{\theta}_K)X_{A_K}X^T \end{pmatrix} = (1/n)X_{A_{diag}}^T W X_{rbind},$$

where $X_{\text{rbind}} = 1 \otimes X$ and $X_{(n\times p)}$ is the reference data matrix; $X_{A_{\text{diag}}} = \text{diag}(X_{A_1}, \ldots, X_{A_K})$ and $X_{A_k(n\times d_k)}$ is the reference data matrix for the kth study; $W = \text{diag}(W_1, \ldots, W_K)$, $W_k = \text{diag}(w_{k1}, \ldots, w_{kn})$, $w_{ki} = [a(\phi_k)g'\{g^{-1}(X_i^T\beta)\}]^{-1}h'(X_{A_k,i}^T\hat{\theta}_k)$ for $k = 1, \ldots, K$, $i = 1, \ldots, n$ and i, and X_i^T and $X_{A_k,i}^T$ are the ith rows of X and X_{A_k}, respectively. Similarly, the matrix form of $U_n(\beta)$ is $U_n(\beta) = (1/n)X_{A_{diag}}^T r$, where $r = (r_1, \ldots, r_K)^T$, $r_k = (r_{k1}, \ldots, r_{kn})^T$ and $r_{ki} = \{1/a(\phi_k)\}\{g^{-1}(X_i^T\beta) - g^{-1}(X_{A_k,i}^T\hat{\theta}_{A_k,i})\}h'(X_{A_k,i}^T\hat{\theta}_{A_k,i})$ for each k and i. Thus, the matrix form of $D_n(\beta)$ is

$$D_n(\beta) = (1/n^2)X_{rbind}^T W X_{A_{diag}} C X_{A_{diag}}^T r. \tag{9}$$

Next, we write $J_n(\beta)$ in a matrix form. Let $G_n(\beta)$ be partitioned by columns as $G_n(\beta) = (G_{n,1}(\beta), \ldots, G_{n,p}(\beta))$, where $G_{n,j}(\beta)$ is a $d \times 1$ column vector for $j = 1, \ldots, p$. Then,

$$\begin{aligned} J_n(\beta) &= \frac{\partial}{\partial\beta}D_n(\beta) = \frac{\partial}{\partial\beta}G_n^T(\beta)CU_n(\beta) \\ &= \begin{pmatrix} \frac{\partial}{\partial\beta}G_{n,1}^T(\beta)CU_n(\beta) \\ \vdots \\ \frac{\partial}{\partial\beta}G_{n,p}^T(\beta)CU_n(\beta) \end{pmatrix} = G_n^T(\beta)CG_n(\beta) + \begin{pmatrix} U_n^T(\beta)C\frac{\partial}{\partial\beta}G_{n,1}(\beta) \\ \vdots \\ U_n^T(\beta)C\frac{\partial}{\partial\beta}G_{n,p}(\beta) \end{pmatrix}. \end{aligned} \tag{10}$$

Then, the matrix form of the first summand is $(1/n^2)X_{rbind}^T W X_{A_{diag}} C X_{A_{diag}}^T W X_{rbind}$. The jth row of the second summand is $r^T X_{A_{diag}} C \partial G_{n,j}(\beta)/\partial\beta$. Note that

$$\frac{\partial}{\partial\beta}G_{n,j}(\beta) = (1/n)X_{A_{diag}}^T L X_{j_{diag}}^* X_{rbind},$$

where $L = \text{diag}(L_1, \ldots, L_K)$, $L_k = \text{diag}(l_{k1}, \ldots, l_{kn})$ and, for each k and i,

$$l_{ki} = -g''\{g^{-1}(X_i^T\beta)\}/(a(\phi_k)[g'\{g^{-1}(X_i^T\beta)\}]^3 h'(X_{A_k,i}^T\hat{\theta}_k));$$

$X_{j_{diag}}^* = \text{diag}(X_{j_{diag}}, \ldots, X_{j_{diag}})$ with K diagonal blocks and $X_{j_{diag}} = \text{diag}(X_{1j}, \ldots, X_{nj})$ for $j = 1, \ldots, p$. Then, for each j, the matrix form of $U_n^T(\beta)C\partial G_{n,j}(\beta)/\partial\beta$ is

$$(1/n^2)r^T X_{A_{diag}} C X_{A_{diag}}^T L X_{j_{diag}}^* X_{rbind}.$$

Then, the second summand of (10) can be rewritten as $(1/n^2)X_{rbind}^T V X_{rbind}$, where $V = diag(v_1, \ldots, v_{nK})$ and v_i is the ith element of the row vector $r^T X_{A_{diag}} C X_{A_{diag}}^T L$. Thus,

$$J_n(\beta) = (1/n^2)X_{rbind}^T(W X_{A_{diag}} C X_{A_{diag}}^T W \;+\; V)X_{rbind} = (1/n^2)X_{\text{rbind}}^T W^* X_{\text{rbind}}. \tag{11}$$

where $W^* = W X_{A_{diag}} C X_{A_{diag}}^T W + V$.

Therefore, plugging (9) and (11) in (8), we get the following tth iteration step

$$\beta^{(t+1)} = \beta^{(t)} - (X_{\text{rbind}}^T W^* X_{\text{rbind}})^{-1} X_{\text{rbind}}^T W X_{A_{\text{diag}}} C X_{A_{\text{diag}}}^T r,$$

which can be seen as the tth step of an iteratively reweighted least squares algorithm.

2.2. *Case II : ϕ and ϕ_k's are unknown.*

When ϕ and ϕ_k's are unknown, we propose to first obtain the GENMETA estimator $\hat{\beta}$ of $\beta^\star$ as above with $\phi_k' s$ replaced by $\hat{\phi}_k$'s. Next, let us consider the estimation of $\phi^\star$, the true value of ϕ. For the kth reduced model, we have an additional score function with respect to ϕ_k, which is

$$s_k(y \mid x_{A_k}; \theta_k, \phi_k) = -\frac{a'(\phi_k)}{a^2(\phi_k)}(yh(x_{A_k}^T\theta_k) - b\{h(x_{A_k}^T\theta_k)\}) + c'(y; \phi_k),$$

where $c'(y; \phi_k)$ is the derivative of $c(y; \phi_k)$ with respect to ϕ_k. Then, we obtain

$$u_k(X; \beta, \phi, \theta_k, \phi_k) = -\frac{a'(\phi_k)}{a^2(\phi_k)}(g^{-1}(X^T\beta)h(X_{A_k}^T\theta_k) - b\{h(X_{A_k}^T\theta_k)\}) + q_k(X; \beta, \phi, \phi_k),$$

where $q_k = E_{Y|X}(c'(Y, \phi_k))$. The distribution of $Y \mid X$ depends on β and ϕ so that q_k also depends on them. Then, the empirical moment vector for ϕ is

$$U_n(\phi) = P_n(u_1(X; \hat{\beta}, \phi, \hat{\theta}_1, \hat{\phi}_1)^T, \ldots, u_K(X; \hat{\beta}, \phi, \hat{\theta}_K, \hat{\phi}_K)^T)^T.$$

We propose to estimate $\phi^\star$ in the GMM framework. Thus, we need to compute the minimizer of $U_n(\phi)^T C U_n(\phi)$, where C is a known weighting matrix. As before, we use the Newton-Raphson's method and it can be written as

$$\phi^{(t+1)} = \phi^{(t)} - J_n^{-1}(\phi^{(t)}) D_n(\phi^{(t)}), \tag{12}$$

where

$$J_n(\phi) = U_n^T(\phi) C \frac{d^2}{d\phi^2} q_n(\phi) + (\frac{d}{d\phi} q_n(\phi))^T C \frac{d}{d\phi} q_n(\phi),$$

$D_n(\phi) = U_n^T(\phi^{(t)}) C dq_n(\phi)/d\phi$ and $q_n(\phi) = P_n(q_1(X; \hat{\beta}, \phi, \hat{\phi}_1), \ldots, q_K(X; \hat{\beta}, \phi, \hat{\phi}_K))^T$.

Thus, when ϕ and ϕ_k's are unknown, we first choose initial estimates $\beta^{(0)}$ and $\phi^{(0)}$. Then, we get the GENMETA estimator $\hat{\beta}$ by using equation (8) until a stopping rule is reached. Subsequently, $\phi^{(0)}$, $\hat{\beta}$ and the study estimates are plugged in equation (12) and the process is repeated until a stopping rule is reached to get the GENMETA estimator of ϕ^*. In each Newton-Raphson's step, the weighting matrix C is estimated by the estimates from the previous step.

If the estimates of the study dispersion parameters, ϕ_k's, are not provided directly, but the the outcomes are standardized ($\text{var}(Y) = 1$), we can obtain them through the following relation based on conditional variance formula

$$a(\hat{\phi}_k) = \frac{1 - (P_n g^{-1}(X_{A_k}^T\hat{\theta}_k)^2 - \{P_n g^{-1}(X_{A_k}^T\hat{\theta}_k)\}^2)}{P_n b''\{h(X_{A_k}^T\hat{\theta}_k)\}},$$

where $h(\cdot) = b'^{-1}(g^{-1}(\cdot))$ and P_n is the empirical measure with the reference data. For normal family where the canonical link is an identity function, we have $b''(\psi) = 1$, which implies the denominator is 1.

3. Full Proof of Theorem 1 and Checking Regularity Assumptions in Two Examples

We first provide a complete proof of Theorem 1 and then check the assumptions for logistic and linear regression models.

Proof of Theorem 1. First, we show the consistency of $\hat{\beta}$. Denote $\hat{\theta}$ and θ^* as stacked vectors of $\hat{\theta}_k$'s and θ_k^*'s, respectively. Denote $U_0(\beta,\theta) = E(U(X;\beta,\theta))$ and $Q_0(\beta) = U_0(\beta,\theta^*)^T C U_0(\beta,\theta^*)$.

By (A1) and Lemma 2.3 of Newey & McFadden (1994), $Q_0(\beta)$ is uniquely minimized at β^*.

By (A2), (A3), (A4) and Lemma 2.4 of Newey & McFadden (1994), $U_0(\beta,\theta)$ is continuous and $U_n(\beta,\theta)$ converges uniformly to $U_0(\beta,\theta)$ for $(\beta,\theta) \in D_\beta \times N_c(\theta^*)$, where $N_c(\theta^*)$ is a compact subset of $N(\theta^*)$ including θ^*. Note that $\hat{\theta}$ is a consistent estimator of θ^*. With probability going to one (wpg1),

$$\sup_{\beta \in D_\beta} ||U_n(\beta,\hat{\theta}) - U_0(\beta,\hat{\theta})|| \leq \sup_{(\beta,\theta)\in D_\beta \times N_c(\theta^*)} ||U_n(\beta,\theta) - U_0(\beta,\theta)||.$$

Then, $U_n(\beta,\hat{\theta}) - U_0(\beta,\hat{\theta})$ converges uniformly in probability to 0 for $\beta \in D_\beta$.

For any $r > 0$, wpg1,

$$\sup_{\beta \in D_\beta} ||U_0(\beta,\hat{\theta}) - U_0(\beta,\theta^*)|| \leq \sup_{\beta \in D_\beta} E(\sup_{||\theta-\theta^*||<r} ||U(\beta,\theta) - U(\beta,\theta^*)||).$$

By (A3), (A4) and dominant convergence theorem, $E(\sup_{||\theta-\theta^*||<r} ||U(\beta,\theta) - U(\beta,\theta^*)||)$ converges to 0 for every $\beta \in D_\beta$ as r decreases to 0. Note that $E(\sup_{||\theta-\theta^*||<r} ||U(\beta,\theta) - U(\beta,\theta^*)||)$ decreases as r decreases for each β. By (A2) and Dini's theorem (see, for example, Theorem 7.13 of Rudin (1976)), $E(\sup_{||\theta-\theta^*||<r} ||U(\beta,\theta) - U(\beta,\theta^*)||)$ converges uniformly in probability to 0 for $\beta \in D_\beta$ as r decreases to 0. Then, $U_0(\beta,\hat{\theta}) - U_0(\beta,\theta^*)$ converges uniformly in probability to 0 for $\beta \in D_\beta$.

By combining the above two results, it follows that $U_n(\beta,\hat{\theta})$ converges uniformly in probability to $U_0(\beta,\theta^*)$ for $\beta \in D_\beta$.

By the triangle and Cauchy-Schwartz inequalities,

$$\begin{aligned}\sup_{\beta \in D_\beta} |Q_n(\beta) - Q_0(\beta)| \leq{}& ||\hat{C}|| \sup_{\beta \in D_\beta} ||U_n(\beta,\hat{\theta}) - U_0(\beta,\theta^*)||^2 \\ &+ 2||\hat{C}|| \sup_{\beta \in D_\beta} ||U_0(\beta,\theta^*)|| \sup_{\beta \in D_\beta} ||U_n(\beta,\hat{\theta}) - U_0(\beta,\theta^*)|| \\ &+ ||\hat{C} - C|| \sup_{\beta \in D_\beta} ||U_0(\beta,\theta^*)||^2\end{aligned}$$

Since $\hat{C}$ is a consistent estimator of C, $||\hat{C}||$ converges in probability to $||C||$, which is finite; $||\hat{C} - C||$ converges in probability to 0. Since $U_0(\beta,\theta^*)$ is continuous for $\beta \in D_\beta$ and D_β is compact, $\sup_{\beta \in D_\beta} ||U_0(\beta,\theta^*)||^2$ is finite. Since $\sup_{\beta \in D_\beta} ||U_n(\beta,\hat{\theta}) - U_0(\beta,\theta^*)||$ converges in probability to 0, $\sup_{\beta \in D_\beta} ||U_n(\beta,\hat{\theta}) - U_0(\beta,\theta^*)||^2$ converges in probability to 0. Thus, $Q_n(\beta) - Q_0(\beta)$ converges uniformly in probability to 0 for $\beta \in D_\beta$. Recall that β^* is the unique minimizer of $Q_0(\beta)$. By Theorem 2.1 of Newey & McFadden (1994), $\hat{\beta}$ is a consistent estimator of β^*.

Next, we derive the asymptotic distribution of the GENMETA estimator $\hat{\beta}$. Note that $\hat{\beta}$ is a solution to

$$G_n(\beta,\hat{\theta})^T \hat{C} U_n(\beta,\hat{\theta}) = 0,$$

where $G_n(\beta,\hat{\theta}) = \partial U_n(\beta,\hat{\theta})/\partial\beta$, the Jacobian of $U_n(\beta,\hat{\theta})$. On the other hand, by mean value theorem,

$$U_n(\hat{\beta},\hat{\theta}) = U_n(\beta^*,\hat{\theta}) + G_n(\bar{\beta},\hat{\theta})(\hat{\beta}-\beta^*),$$

where $\bar{\beta}$ denotes a matrix each column of which corresponds to each element of $U_n(\beta,\hat{\theta})$. After left multiplying $G_n(\hat{\beta},\hat{\theta})^T\hat{C}$ to the above identity, it follows

$$n^{1/2}(\hat{\beta}-\beta^*) = -M_n n^{1/2}U_n(\beta^*,\hat{\theta}),$$

where $M_n = (G_n(\hat{\beta},\hat{\theta})^T\hat{C}G_n(\bar{\beta},\hat{\theta}))^{-1}G_n(\hat{\beta},\hat{\theta})^T\hat{C}$.

Consider M_n. Since $\hat{\beta}$ is a consistent estimator of β^*, each column of $\bar{\beta}$ is a consistent estimator of β^*. On the other hand, $\hat{\theta}$ is a consistent estimator of θ^*. By (A5), (A6) and Lemma 2.4 of Newey & McFadden (1994), $G_n(\beta,\theta)$ converge uniformly to continuous $E\{\partial U(X;\beta,\theta)/\partial\beta\}$ for $(\beta,\theta)\in D_\beta\times N_c(\theta^*)$, where $N_c(\theta^*)$ is a compact subset of $N(\theta^*)$, including θ^*. Since $\hat{\beta}$ and each column of $\bar{\beta}$ converge in probability to β^* and $\hat{\theta}$ is a consistent estimator of θ^*, by, for example, Theorem 9.4 of Keener (2010), both $G_n(\hat{\beta},\hat{\theta})$ and $G_n(\bar{\beta},\hat{\theta})$ converges in probability to $\Gamma = E\{\partial U(X;\beta^*,\theta^*)/\partial\beta\}$. Thus, by noting $\hat{C}\to C$ in probability, M_n converges in probability to $(\Gamma^T C\Gamma)^{-1}\Gamma^T C$.

Consider $n^{1/2}U_n(\beta^*,\hat{\theta})$. By mean value theorem,

$$U_n(\beta^*,\hat{\theta}) = U_n(\beta^*,\theta^*) + V_n(\beta^*,\bar{\theta})(\hat{\theta}-\theta^*),$$

where V_n is the Jacobian of $U_n(\beta^*,\theta)$ as a function of θ and $\bar{\theta}$ is a matrix each column of which corresponds to each element of $U_n(\beta^*,\theta)$. Thus,

$$n^{1/2}U_n(\beta^*,\hat{\theta}) = n^{1/2}U_n(\beta^*,\theta^*) + V_n(\beta^*,\bar{\theta})n^{1/2}(\hat{\theta}-\theta^*).$$

By (A9) and central limit theorem, $n^{1/2}U_n(\beta^*,\theta^*)\xrightarrow{d}N(0,\Delta)$. Since $\hat{\theta}$ is a consistent estimator of θ^*. each column of $\bar{\theta}$ converges in probability to θ^*. Similar to the above argument, by (A7), (A8), Lemma 2.4 of Newey & McFadden (1994) and Theorem 9.4 of Keener (2010),

$$V_n(\beta^*,\bar{\theta})\to\text{diag}(W_1,W_2,\ldots,W_K) \quad \text{in probability},$$

where, for $k=1,2,\ldots,K$, $W_k = E\{\partial u_k(X,\beta^*,\theta_k)/\partial\theta_k\}\mid_{\theta_k=\theta_k^*}$. The K study data sets are independent. So are $\hat{\theta}_k$'s. Note that $n_k/n\to c_k$, where c_k is a positive constant for $k=1,2,\ldots,K$. Then $n^{1/2}(\hat{\theta}-\theta^*)$ converges in distribution to

$$N(0,\text{diag}((1/c_1)\Sigma_1,(1/c_2)\Sigma_2,\ldots,(1/c_K)\Sigma_K)).$$

Since the K data sets and the reference data are independent, the above results imply that $n^{1/2}U_n(\beta^*,\hat{\theta})$ converges in distribution to $N(0,\Delta+\Lambda)$, where Λ is a block diagonal matrix whose kth block is $(1/c_k)W_k\Sigma_k W_k^T$ for $k=1,\ldots,K$.

Therefore, with the above two results on M_n and $n^{1/2}U_n(\beta^*,\hat{\theta})$ and by Slutsky's theorem, the asymptotic normality of $n^{1/2}(\hat{\beta}-\beta^*)$ follows. □

Example 1 *(Check Assumptions for Logistic Regression Model).* Suppose the maximal model is

$$Y\mid X\sim\text{Bernoulli}\Big\{\frac{1}{1+\exp(-X^T\beta^*)}\Big\},$$

where $X = (1, X^T)^T$, $X = (X_1, \ldots, X_d)^T$ is the vector of covariates and $\beta^* = (\beta_0^*, \beta_1^*, \ldots, \beta_p^*)^T$ is the vector of coefficients of interest. There are K independent studies and the reduced model of the kth study is

$$Y \mid X_{A_k} \sim \text{Bernoulli}\Big\{\frac{1}{1+\exp(-X_{A_k}^T\theta_k)}\Big\},$$

where $X_{A_k} = (1, X_{A_k}^T)^T$, X_{A_k} is a sub-vector of X with $A \subset \{1, 2, \ldots, p\}$. For example, $X_A = (X_1, X_2)^T$ when $A = \{1, 2\}$.

The global identification assumption (A1) usually holds and D_β is a compact set. Next, we check the assumptions (A3) to (A9). The moment functions from the kth study is

$$u_k(X; \beta, \theta_k) = \Big(\frac{1}{1+e^{-X^T\beta}} - \frac{1}{1+e^{-X_{A_k}^T\theta_k}}\Big)X_{A_k}.$$

It is a continuous function of β and θ_k. Then, (A3) is satisfied. Note that

$$\sup_{(\beta,\theta)\in D_\beta\times N(\theta^*)} ||\Big(\frac{1}{1+e^{-X^T\beta}} - \frac{1}{1+e^{-X_{A_k}^T\theta_k}}\Big)X_{A_k}|| \leq 2||X||_1,$$

where $||\cdot||$ and $||\cdot||_1$ are the l_2 and l_1 norms, respectively. Then, given $E(|X_i|) < \infty$ for each i, (A4) is satisfied. Also,

$$\frac{\partial}{\partial\beta}u_k(X; \beta, \theta_k) = \frac{e^{-X^T\beta}}{(1+e^{-X^T\beta})^2}X_{A_k}X^T, \tag{13}$$

which does not depend on θ_k and is continuous for each β. Then, (A5) is verified. Note that

$$\sup_{(\beta,\theta)\in D_\beta\times N(\theta^*)} ||\frac{e^{-X^T\beta}}{(1+e^{-X^T\beta})^2}X_{A_k}X^T|| \leq ||XX^T||_1.$$

Given $E(X_i^2) < \infty$ for each i, (A6) is satisfied. Note that

$$\frac{\partial}{\partial\theta_k}u_k(X; \beta^*, \theta_k) = -\frac{e^{-X_{A_k}^T\theta_k}}{(1+e^{-X_{A_k}^T\theta_k})^2}X_{A_k}X_{A_k}^T,$$

which is continuous for each θ_k. Then, (A7) is satisfied. Note that

$$\sup_{(\beta,\theta)\in D_\beta\times N(\theta^*)} ||-\frac{e^{-X_{A_k}^T\theta_k}}{(1+e^{-X_{A_k}^T\theta_k})^2}X_{A_k}X_{A_k}^T|| \leq ||XX^T||_1.$$

Given $E(X_i^2) < \infty$ for each i, (A8) is satisfied. The absolute value of each element of $\Delta(\beta^*, \theta^*)$ is less than 1, $E(|X_i|)$ or $E(|X_iX_j|)$ for each i and j. Given $E(X_i^2) < \infty$, $\Delta(\beta^*, \theta^*)$ is finite. Note that $\Gamma(\beta^*, \theta_k^*)$ is a stacked matrix of (13) for $k = 1, \ldots, K$. Given each covariate of the maximal model is in at least one reduced model and $E[\{e^{-X^T\beta}/(1+e^{-X^T\beta})^2\}XX^T]$ is positive definite, $\Gamma(\beta^*, \theta^*)$ is of full rank. Then, (A9) is verified.

Example 2 *(Check Assumptions for Linear Regression Model).* Suppose the true maximal model is

$$Y \mid X \sim N(X^T\beta^*, \sigma^{*2}),$$

where $X = (X_1, X_2, \ldots, X_p)^T$; $\beta^* = (\beta_1^*, \beta_2^*, \ldots, \beta_p^*)^T$; $E(X) = 0$ and $E(Y) = 0$, that is, both X and Y are centered. There are K independent studies and the reduced model of the

kth study is

$$Y \mid X_{A_k} \sim N(X_{A_k}^T \theta_k, \sigma_k^2).$$

For simplicity, assume σ^{*2} is known and the unknown parameter is β^*. The case with unknown σ^{*2} can be similarly considered.

The moment functions from the kth reduced model is

$$u_k(X; \beta; \theta_k, \sigma_k^2) = \frac{1}{\sigma_k^2}(X_{A_k} X^T \beta - X_{A_k} X_{A_k}^T \theta_k),$$

which is linear in β. Note that

$$\frac{\partial}{\partial \beta} u_k(X; \beta; \theta_k, \sigma_k^2) = \frac{1}{\sigma_k^2} X_{A_k} X^T. \tag{14}$$

Given each covariate of the maximal model is in at least one reduced model and $E(XX^T)$ is positive definite, $\Gamma(\beta^*, \{\theta_k^*\}, \{\sigma_k^{*2}\}) = \partial u_k(X; \beta^*; \{\theta_k^*\}, \{\sigma_k^{*2}\})/\partial \beta$ is of full rank. Given C is positive definite, (A1) is satisfied. Suppose D_β is a compact set. Then, (A2) is satisfied.

Next, we check the assumptions (A3) to (A9). Note that $u_k(X; \beta; \theta_k, \sigma_k^2)$ is continuous for every $(\beta, \theta_k, \sigma_k^2)$. Then, (A3) is satisfied. Note that

$$\sup_{(\beta, \theta_k, \sigma_k^2)} ||\frac{1}{\sigma_k^2}(X_{A_k} X^T \beta - X_{A_k} X_{A_k}^T \theta_k)|| \leq \frac{1}{\sigma_k^2}(||\beta|| + ||\theta_k||)||XX^T||_1,$$

Denote a finite upper bound of $||\beta||$ for $\beta \in D_\beta$ as $C(\beta)$, a finite upper bound of $||\theta_k||$ for $\theta_k \in N(\theta_k^*)$ as $C(\theta_k)$, and a positive finite lower bound of σ_k^2 for $\sigma_k^2 \in N(\theta_k^*)$ as σ_L^2. The supremum of $(1/\sigma_k^2)(||\beta|| + ||\theta_k||)$ for $(\beta, \theta_k, \sigma_k^2) \in D_\beta \times N(\theta_k^*) \times N(\sigma_k^{*2})$ is bounded by $(1/\sigma_L^2)(C(\beta) + C(\theta_k))$. Given $E(X_i^2) < \infty$ for each i, (A4) is satisfied. Note that $\partial u_k(X; \beta; \theta_k, \sigma_k^2)/\partial \beta$ does not depend on β and θ_k and is continuous for each σ_k^2. Then, (A5) is satisfied. Note that

$$\sup_{\sigma_k^2 \in N(\sigma_k^{*2})} ||\frac{1}{\sigma_k^2} X_{A_k} X^T|| \leq \frac{1}{\sigma_L^2}||XX^T||_1.$$

Given $E(X_i^2) < \infty$ for each i, (A6) is satisfied. Note that

$$\frac{\partial}{\partial(\theta_k, \sigma_k^2)} u_k(X; \beta; \theta_k, \sigma_k^2) = \{-\frac{1}{\sigma_k^2} X_{A_k} X_{A_k}^T, -\frac{1}{\sigma_k^4}(X_{A_k} X^T \beta - X_{A_k} X_{A_k}^T \theta_k)\},$$

which is continuous for every $(\beta, \theta_k, \sigma_k^2)$. Then, (A7) is satisfied. For every $(\beta, \theta_k, \sigma_k^2) \in D_\beta \times N(\theta_k^*, N(\sigma_k^{*2}))$, the l_2 norm of the above partial derivative is less than or equal to

$$\frac{1}{\sigma_L^2} + \frac{1}{\sigma_L^4}(C(\beta) + C(\theta_k))||XX^T||_1.$$

Given $E(X_i^2) < \infty$ for each i, (A8) is satisfied. Each element of $\Delta(\beta^*, \{\theta_k^*\}, \{\sigma_k^{*2}\})$ is equal to a constant times $E(X_{i_1} X_{i_2} X_{i_3} X_{i_4})$ for some i_1, i_2, i_3, i_4. Given $E(X_i^4) < \infty$ for each i, Δ is finite. Note that $\Gamma(\beta^*, \{\theta_k^*\}, \{\sigma_k^{*2}\})$ is a stacked matrix of (14) for $k = 1, \ldots, K$. As in checking (A2), given each covariate of the maximal model is in at least one reduced model and $E(XX^T)$ is positive definite, Γ is of full rank. Then, (A9) is verified.

4. SIMULATION RESULTS FOR LOG-NORMALLY DISTRIBUTED COVARIATES

Table 1: Robustness of GENMETA Estimation (Log-normally Distributed Covariates)

Setting	Study-I	Study-II	Study-III	Reference	β_i^*	Bias	SD (ESD)	RMSE	CR	AL
	μ_b	μ_b	μ_b	μ_b	β_1^*	.010	.076 (.075)	.077	.941	.288
I	σ_b^2	σ_b^2	σ_b^2	σ_b^2	β_2^*	.011	.064 (.061)	.065	.947	.237
	ρ_b	ρ_b	ρ_b	ρ_b	β_3^*	.006	.066 (.064)	.066	.954	.246
	μ_b	μ_h	μ_m	μ_b	β_1^*	.010	.079 (.072)	.079	.930	.272
	σ_b^2	σ_b^2	σ_b^2	σ_b^2	β_2^*	.002	.056 (.054)	.056	.948	.211
	ρ_b	ρ_b	ρ_b	ρ_b	β_3^*	-.002	.062 (.058)	.062	.945	.222
	μ_b	μ_b	μ_b	μ_b	β_1^*	.032	.088 (.088)	.094	.930	.339
II	σ_b^2	σ_h^2	σ_l^2	σ_b^2	β_2^*	-.002	.062 (.057)	.062	.941	.221
	ρ_b	ρ_b	ρ_b	ρ_b	β_3^*	-.005	.074 (.074)	.074	.967	.286
	μ_b	μ_h	μ_m	μ_b	β_1^*	.021	.079 (.077)	.081	.929	.294
	σ_b^2	σ_h^2	σ_l^2	σ_b^2	β_2^*	.0005	.055 (.055)	.055	.956	.213
	ρ_b	ρ_b	ρ_b	ρ_b	β_3^*	-.008	.065 (.064)	.065	.954	.246
	μ_b	μ_b	μ_b	μ_b	β_1^*	-.062	.107 (.118)	.124	.934	.382
	σ_b^2	σ_b^2	σ_b^2	σ_b^2	β_2^*	.021	.070 (.065)	.073	.930	.250
	ρ_b	ρ_b	ρ_b	ρ_h	β_3^*	.030	.087 (.096)	.092	.956	.322
	μ_b	μ_b	μ_b	μ_b	β_1^*	.039	.072 (.069)	.081	.891	.264
	σ_b^2	σ_b^2	σ_b^2	σ_b^2	β_2^*	.023	.065 (.062)	.069	.932	.240
	ρ_b	ρ_b	ρ_b	ρ_l	β_3^*	.018	.061 (.058)	.064	.930	.224
	μ_b	μ_b	μ_b	μ_b	β_1^*	.053	.079 (.075)	.095	.866	.290
III	σ_b^2	σ_b^2	σ_b^2	σ_b^2	β_2^*	.019	.065 (.063)	.067	.942	.242
	ρ_l	ρ_b	ρ_h	ρ_l	β_3^*	.012	.068 (.064)	.069	.935	.249
	μ_b	μ_b	μ_b	μ_b	β_1^*	.032	.089 (.084)	.095	.912	.322
	σ_b^2	σ_b^2	σ_b^2	σ_b^2	β_2^*	.010	.062 (.062)	.063	.946	.240
	ρ_l	ρ_b	ρ_h	ρ_b	β_3^*	-.009	.073 (.071)	.073	.942	.273
	μ_b	μ_b	μ_b	μ_b	β_1^*	-.025	.113 (.108)	.116	.954	.407
	σ_b^2	σ_b^2	σ_b^2	σ_b^2	β_2^*	.017	.065 (.064)	.067	.951	.248
	ρ_l	ρ_b	ρ_h	ρ_h	β_3^*	-.002	.091 (.091)	.091	.965	.347
			μ_b	μ_b	β_1^*	.007	.096 (.104)	.096	.968	.365
IV	$X_1 > -0.5$,	$X_2 > 0$	σ_b^2	σ_b^2	β_2^*	.242	.353 (.117)	.428	.572	.401
	$X_2 < 0.5$		ρ_b	ρ_b	β_3^*	-.015	.067 (.081)	.068	.971	.283

Biases, standard deviation (SD), estimated standard deviation (ESD), square roots of mean square errors (RMSE), coverage rates (CR), and average lengths (AL) of 95% confidence intervals of the GENMETA estimates using the study covariance estimators in the setting of logistic regression. In setting (I), data are simulated in ideal setting where the covariate distribution is a log-normal distribution with the natural logarithm of the covariates being characterized by mean, sd and correlation of normal variates and are assumed to same across all populations. In setting (II)-(IV), the assumption is violated by creating variations in mean/sd, correlations of the underlying normal distribution and selection criterion across the studies and reference sample. The vector of means, variances and correlations of the underlying normal covariates are denoted by $\mu_* = (\mu_1, \mu_2, \mu_3)$, $\sigma_*^2 = (\sigma_1^2, \sigma_2^2, \sigma_3^2)$ and $\rho_* = (\rho_{12}, \rho_{23}, \rho_{13})$ for $* \in \{b, l, m, h\}$, where $\mu_b = (0, 0, 0)$, $\mu_m = (0.5, 0.5, 0.5)$, $\mu_h = (1, 1, 1)$; $\sigma_b^2 = (1, 1, 1)$, $\sigma_l^2 = (0.5, 0.5, 0.5)$, $\sigma_h^2 = (2, 2, 2)$ and $\rho_b = (0.3, 0.6, 0.1)$, $\rho_h = (0.4, 0.8, 0.2)$, $\rho_l = (0.2, 0.4, 0)$. Estimated standard deviation are obtained by the asymptotic formula (2) in the main paper and used to construct 95% confidence interval.

REFERENCES

Keener, R. W. (2010). *Theoretical Statistics: Topics for a Core Course*. Springer.

McCullagh, P. & Nelder, J. A. (1989). *Generalized Linear Models*. Chapman & Hall/CRC, 2nd ed.

Newey, W. K. & McFadden, D. (1994). Chapter 36 large sample estimation and hypothesis testing. In *Handbook of Econometrics*, vol. Volume 4. Elsevier, pp. 2111–2245.

Rudin, W. (1976). *Principles of Mathematical Analysis*. McGraw Hill, 3rd ed.

[*Received on* x xx 2017. *Editorial decision on* x xx 2017]